Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Linear Algebra and its Applications 413 (2006) 327–341
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

A note on local and global convergence analysis of iterative aggregation–disaggregation methods

Ivo Marek [1], Ivana Pultarová *

Department of Mathematics, Faculty of Civil Engineering, Czech Technical University in Prague, Thákurova 7, 166 29 Prague, Czech Republic

Received 9 November 2004; accepted 24 July 2005
Available online 12 September 2005
Submitted by B. Shader

Abstract

The purpose of the paper is to present some convergence properties of the iterative aggregation–disaggregation method for computing a stationary probability distribution vector of a column stochastic matrix. A sufficient condition for the local convergence property and the corresponding rate of convergence are established. Some global convergence considerations are presented. Several illustrative examples are included.

AMS classification: 65F10; 15A51; 60J10

Keywords: Stochastic matrix; Stationary probability distribution vector; Iterative aggregation–disaggregation method

1. Introduction

The iterative aggregation–disaggregation (IAD) method presented in this paper belongs to a class of multilevel methods for solving linear systems [2,6,12,13]. We

* Corresponding author. Tel.: +420 22435 4408; fax: +420 233 332 732.
E-mail addresses: marek@mbox.ms.mff.cuni.cz, marekivo@mat.fsv.cvut.cz (I. Marek), ivana@mat.fsv.cvut.cz (I. Pultarová).
[1] Tel.: +420 22435 4381.

0024-3795/$ - see front matter
doi:10.1016/j.laa.2005.08.001

use this approach to obtain the stationary probability distribution vector of a stochastic matrix, i.e. the aim is to find a vector $\hat{x}$ such that

$$B\hat{x} = \hat{x},$$

where B is a column stochastic matrix and $\hat{x}$ is appropriately normalized. The typical characteristic within the class mentioned is that the corresponding basic iteration matrix is a polynomial in B. This property allows to analyze this type of IAD methods in more detail. In particular, we show that we are able to control the number of relaxations on the fine level (Theorem 1).

We present some results on convergence analysis of IAD methods. After introducing notations, in the second part, we present sufficient conditions for local convergence for one of the simplest form of the IAD method and we derive the asymptotic rate of convergence for this case. We show that when B contains at least one strictly positive row then there exists a neighborhood of $\hat{x}$ such that for every initial approximation in it, the IAD method yields a sequence which converges to $\hat{x}$. The result is compared with an iterative IAD method which uses partitioning into two groups and thus it can be analysed deeper [5]. Two examples show that the divergence of the IAD method might be explained in some cases similar to the two groups case. Our analysis shows that the iteration process investigated in [5] can be treated as a particular case of the scheme shown in [10,11]. This explains the appearance of some of its additional properties if the special method are compared with the more general IAD methods. On the other hand, some of the tools applied in [5] can be utilized to analyzing more general situations as we show on two examples demonstrating possible divergence of a class of IAD methods with the power method with B as the basic iteration. The second result concerns the global convergence. In Section 4, a sufficient condition for the global convergence property of IAD method is shown. In the end of the paper, the relation to previous results and to some other IAD methods are discussed.

1.1. Definitions and basic relations

Let B be an $N \times N$ column stochastic matrix, irreducible and not cyclic (for definitions see e.g. [1,13]). Then according to [1,13,15], matrix B possesses a simple eigenvalue 1 and the other eigenvalues are smaller than 1 in the modulus. The eigenvector $\hat{x}$ corresponding to the eigenvalue 1 is strictly positive and it is called *the stationary probability distribution vector* of matrix B.

Let Q and Z denote the spectral decomposition of B, fulfilling $B = Q + Z$, $Q^2 = Q$, $QZ = ZQ = 0$ and $\lim_{k\to\infty} Z^k = 0$. Matrix Q is the projection matrix corresponding to the eigenvalue 1. Note that $B\hat{x} = \hat{x} = Q\hat{x}$.

In this paper, $||\cdot||$ denotes the 1-norm, $||M|| = \max_{j=1,\ldots,N} \sum_{i=1}^{N} |M_{ij}|$, $r(M)$ denotes the spectral radius of a matrix M [4] and e is the vector of ones, $e = (1, \ldots, 1)^{\mathrm{T}}$, the size of the vector can change according to the context.

Now we introduce the IAD method. Let $G_1, \ldots, G_n$, $n \leqslant N$, be the *aggregation groups* of events numbered with $1, 2, \ldots, N$. The sets G_i, $i = 1, \ldots, n$, are considered to be disjoint and $\bigcup_{i=1}^{n} G_i = \{1, 2, \ldots, N\}$.

Let us define the *restriction* (*aggregation*) $n \times N$ matrix R, $R_{ij} = 1$ if $j \in G_i$ and $R_{ij} = 0$ otherwise. For any positive x the *prolongation* (*disaggregation*) $N \times n$ matrix $S(x)$ is defined by

$$S(x)_{ij} = \frac{x_i}{\sum_{k \in G_j} x_k}$$

if $i \in G_j$ and $S(x)_{ij} = 0$ otherwise. Let $P(x)$ be the projection matrix given by $P(x) = S(x)R$. Note that $RS(x) = I$, I is the identity matrix. Finally let $T = M^{-1}W$ be a matrix arising from some splitting of $I - B$ which is of weak nonnegative type, i.e. $M^{-1}W \geqslant 0$ [1].

We will use the following relations. When z is a positive $n \times 1$ vector, $||z|| = 1$, then $S(x)z$ is a positive $N \times 1$ vector satisfying $||S(x)z|| = 1$. Multiplying any nonnegative vector x, $||x|| = 1$, by matrix Q, we obtain $\hat{x}$, which is equal to any column of Q. Then also $QS(x)z = \hat{x}$ and $QP(x) = Q$. On the other hand, $P(x)Q \neq Q$ in general, but $P(\hat{x})Q = Q$.

The IAD method consists of several repeating steps. A description of the IAD algorithm considered in this paper follows. Here, the upper vector index denotes an order of the vector in a sequence, while the upper matrix index is an exponent.

Algorithm 1 (*IAD method*)

Step 1. An elementwise positive initial approximation x^0, $||x^0|| = 1$, is selected. The value of k is set to 0.

Step 2. A positive integer s is chosen and the $n \times n$ *aggregated* matrix

$$RB^sS(x^k)$$

is constructed. The associated problem is solved, i.e. the vector z is found, which fulfilles

$$RB^sS(x^k)z = z$$

$||z|| = 1$. This step can be called *the solution on the coarse level*.

Step 3. The prolonged vector $x^{k+1,1}$ of the original size N is computed by

$$x^{k+1,1} = S(x^k)z.$$

Step 4. The next approximation x^{k+1} is computed by $x^{k+1} = T^t x^{k+1,1}$ for the appropriate positive integer t. This step can be called the *smoothing* step or the *correction on the fine level*.

Step 5. The test for convergence is evaluated and then the algorithm finishes with the approximate solution x^{k+1} or continues with Step 2 and with k increased by 1.

Note that all the computed vectors x^k are positive. For any positive x, the aggregated matrix $RB^sS(x)$ is stochastic, irreducible and it is not cyclic. The latter property is a consequence of the acyclicity of B [10,11]. Computing z in Step 2 is assumed to be carried out exactly. In place of the iteration matrix T one can choose any nonnegative matrix with the properties $T\hat{x} = \hat{x}$ and $I - B = M(I - T)$ with some invertible M. Of course, a more sophisticated correction (Step 4) requires more time, but also achieves faster convergence. For example, M can equal to I, M can be the block diagonal part of $I - B$ or M can equal to the sum of the lower block triangle and the block diagonal of $I - B$, etc. Then Step 4 with one of the three mentioned above smoothing methods is referred as power method, block Jacobi method and block Gauss–Seidel method, respectively. Using the notation from [11], Algorithm 1 can be identified as SPV(B; I, B; $P(x)$; s, T; x^0; ϵ).

We now derive the operator which controls the convergence of Algorithm 1.

Lemma 1. *The approximations x^k given by Algorithm* 1, $k = 0, 1, 2, \ldots$, *follow the formula*

$$x^{k+1} - \hat{x} = J(x^k)(x^k - \hat{x}),$$

where

$$J(x) = T^t(I - P(x)Z^s)^{-1}(I - P(x)).$$

Proof. The proof can be found in [10,11] but for the completeness, we present it here. Note that $I - RZ^sS(x_k)$ is invertible due to the absence of 1 in the spectrum of Z^s and thus in the spectrum of $RZ^sS(x_k)$. More precisely, if it was

$$RZ^sS(x)u = u$$

for some $x > 0$ and nonzero u, then for $v = S(x)u$, we have $v \neq 0$ and

$$P(x)Z^sv = v.$$

Having $QP(x) = Q$ and multiplying this equation by Q from the left, we see that $Qv = 0$. Then we have

$$P(x)Z^sv + P(x)Qv = P(x)B^sv = v,$$

which means that v is an eigenvector of the irreducible stochastic matrix $P(x)B^s$ corresponding to its eigenvalue 1, thus $v > 0$. But $v > 0$ contradicts $Qv = 0$. Then $RZ^sS(x)u = u$ cannot be the case for any nonzero u. Continuing with the derivation of x^{k+1}, we can write

$$\begin{aligned} RB^sS(x^k)z &= z, \\ RZ^sS(x^k)z + RQS(x^k)z &= z, \\ (I - RZ^sS(x_k))^{-1}R\hat{x} &= z. \end{aligned}$$

We have

$$\begin{aligned} x^{k+1,1} &= S(x^k)z = S(x^k)(I - RZ^s S(x^k))^{-1} R\hat{x} \\ &= (I - P(x^k)Z^s)^{-1} P(x^k)\hat{x}. \end{aligned}$$

The approximation error vector of x^{k+1} satisfies

$$x^{k+1} - \hat{x} = T^t S(x^k)z - \hat{x} = T^t (I - P(x^k)Z^s)^{-1} P(x^k)\hat{x} - \hat{x}.$$

Reminding that $T^t\hat{x} = \hat{x}$, $Z\hat{x} = 0$ and $P(x^k)x^k = x^k$, we have

$$\begin{aligned} x^{k+1} - \hat{x} &= T^t (I - P(x^k)Z^s)^{-1} P(x^k)(I - Z^s)\hat{x} - \hat{x} \\ &= T^t (I - P(x^k)Z^s)^{-1} (I - P(x^k)Z^s + P(x^k) - I)\hat{x} - \hat{x} \\ &= T^t (I - P(x^k)Z^s)^{-1} (I - P(x^k))(x^k - \hat{x}). \end{aligned}$$

The proof is complete. □

Lemma 2. *If $T = B$ and $t \geqslant s$ then*

$$J(x) = B^{t-s} K(x)$$

or equivalently

$$J(x) = Z^{t-s} K(x),$$

where $K(x)$ can be expressed recurrently

$$K(x) = B^s (I - P(x) + P(x)K(x))$$

or equivalently

$$K(x) = Z^s (I - P(x) + P(x)K(x)).$$

Proof. We have

$$J(x) = B^{t-s} K(x) = B^{t-s} B^s (I - P(x)Z^s)^{-1} (I - P(x))$$

then

$$\begin{aligned} K(x) &= B^s (I - P(x)Z^s)^{-1} (I - P(x)) \\ &= B^s (I - P(x)) + B^s P(x) Z^s (I - P(x)Z^s)^{-1} (I - P(x)) \\ &= B^s (I - P(x)) + B^s P(x) B^s (I - P(x)Z^s)^{-1} (I - P(x)) \\ &= B^s (I - P(x)) + B^s P(x) K(x). \end{aligned}$$

The both equivalencies in the proposition follow immediately from $B^k = Q + Z^k$ for any positive integer k. This completes the proof. □

2. Local convergence condition

In this section we consider the IAD method with its simplest basic iteration matrix. We consider $T = B$ and $s = t = 1$ in Algorithm 1, up to some cases, where T, s and t will be assign explicitly. We show a sufficient condition for the local convergence, i.e. we find out under what circumstances there exists a neighborhood of $\hat{x}$ that for every x^0 in it, the Algorithm 1 results in a sequence of vectors convergent to $\hat{x}$. In [13], it is shown that the exact solution $\hat{x}$ is the fixed point of the IAD method. But more exact convergence conditions has not yet been established. There are more results available for the case of partitioning into just two groups [5,7,8]. Unfortunately they mostly cannot be used for the n groups case.

Proposition 1. *Let B contain at least one positive row, $B \geqslant be^{\mathrm{T}}$, $b \geqslant 0$, $||b|| = \delta$ and $\delta \in (0, 1)$. Then for $T = B$ and $s = t = 1$, matrix $J(x)$ can be expressed in the form*:

$$J(x) = (B - be^{\mathrm{T}})(I - P(x)(B - be^{\mathrm{T}}))^{-1}(I - P(x)).$$

Proof. The proof is similar to the proof of Lemma 1 then we can proceed faster. We have

$$\begin{aligned} RBS(x^k)z &= z, \\ R(B - be^{\mathrm{T}})S(x^k)z + Rb &= z, \\ (I - R(B - be^{\mathrm{T}})S(x^k))^{-1}Rb &= z \end{aligned}$$

since $e^{\mathrm{T}}z = 1$. The invertibility of $I - R(B - be^{\mathrm{T}})S(x^k)$ follows from $R(B - be^{\mathrm{T}})S(x^k) \geqslant 0$ and $||R(B - be^{\mathrm{T}})S(x^k)|| = 1 - \delta$. Then the error vector of x^{k+1} is

$$\begin{aligned} x^{k+1} - \hat{x} &= BS(x^k)(I - R(B - be^{\mathrm{T}})S(x^k))^{-1}Rb - \hat{x} \\ &= (B - be^{\mathrm{T}})S(x^k)(I - R(B - be^{\mathrm{T}})S(x^k))^{-1}Rb + b - \hat{x} \\ &= (B - be^{\mathrm{T}})P(x^k)(I - (B - be^{\mathrm{T}})P(x^k))^{-1}b + b - \hat{x} \\ &= (I - (B - be^{\mathrm{T}})P(x^k))^{-1}b - \hat{x} \\ &= (I - (B - be^{\mathrm{T}})P(x^k))^{-1}(b - \hat{x} + (B - be^{\mathrm{T}})P(x^k)\hat{x}) \\ &= (I - (B - be^{\mathrm{T}})P(x^k))^{-1}(B - be^{\mathrm{T}})(P(x^k) - I)\hat{x} \\ &= (B - be^{\mathrm{T}})(I - P(x^k)(B - be^{\mathrm{T}}))^{-1}(I - P(x^k))(x^k - \hat{x}), \end{aligned}$$

which finishes the proof. □

Let us stress, that if B is elementwise positive and $B \geqslant \delta Q$, then

$$J(x) = (B - \delta Q)(I - P(x)(B - \delta Q))^{-1}(I - P(x)).$$

This is a straightforward consequence of Proposition 1, when considering $be^{\mathrm{T}} = \delta Q = \delta \hat{x} e^{\mathrm{T}}$.

Definition 1. For any positive vector v, we denote by $||M||_v$ the value $\min\{\alpha;\ v^{\mathrm{T}}|M| \leqslant \alpha v^{\mathrm{T}}\}$. It can be called *the matrix norm induced by the vector* v.

Note that if matrix M is elementwise nonnegative, $||M||_v \geqslant r(M)$ for any positive v [15, Theorem 2.9]. The verification of this consists in the transformation of M to a proper matrix similar to M and in the application of Gershgorin theorem.

Now we prove the local convergence of the IAD method both for B positive or for B containing at least one positive row. The proof of the following proposition is based on some ideas presented in [9].

Proposition 2. *Let* $T = B$ *and* $s = t = 1$ *in Algorithm* 1.

(a) *When* $B \geqslant \delta Q$ *for some* $\delta \in (0, 1)$, *then the spectral radius of* $J(\hat{x})$ *is less than or equal to* $1 - \delta$.
(b) *When* $B \geqslant be^{\mathrm{T}}$, $b \geqslant 0$, $||b|| = \delta$, $\delta \in (0, 1)$, *then the spectral radius of* $J(\hat{x})$ *is less than or equal to* $\sqrt{1 - \delta}$.

Proof. The beginning of the proof does not differ for the assumptions (a) or (b) respectively. We can denote $A = B - \delta Q$ or $A = B - be^{\mathrm{T}}$. Note that $A \geqslant 0$ and $||A|| = 1 - \delta$ in both cases. For such As and for any positive vector x we have from Proposition 1

$$\begin{aligned} J(x) &= A(I - P(x)A)^{-1}(I - P(x)) \\ &= A(I - P(x)) + AP(x)A(I - P(x)A)^{-1}(I - P(x)) \\ &= A(I - P(x)) + AP(x)J(x). \end{aligned}$$

For $x \geqslant 0$, let $D(x)$ be an $N \times N$ diagonal matrix in which $D(x)_{ii} = \sqrt{x_i}$, $i = 1, \ldots, N$. Denote $P_s(x) = D(x)^{-1}P(x)D(x)$, $J_s(x) = D(x)^{-1}J(x)D(x)$, $A_s(x) = D(x)^{-1}AD(x)$ and $x_s = D(x)^{-1}x$. Note that $x_s = D(x)e$. In this proof, the index s will not denote an element of a vector, but a matrix or a vector obtained after the introduced transformation.

We now show that $P_s(x)$ is symmetric. Supposing $(P_s(x))_{ij} \neq 0$, we have $(P(x))_{ij} \neq 0$. From the notations in Section 2, it follows that the ith and jth events belong to the same aggregation group, say G_k. Then

$$(P(x))_{ij} = \frac{x_i}{\sum_{l \in G_k} x_l}$$

and thus

$$(P_s(x))_{ij} = \frac{\sqrt{x_j}}{\sqrt{x_i}} \frac{x_i}{\sum_{l \in G_k} x_l} = \frac{\sqrt{x_i x_j}}{\sum_{l \in G_k} x_l} = (P_s(x))_{ji}.$$

We see that $P_s(x)$ is symmetric, thus it is an orthogonal projection [14].

For a proper x we estimate an upper bound for $||A_s(x)||_2^2$, where $||\cdot||_2$ is the spectral norm, $||M||_2 = r(M^{\mathrm{T}}M)^{1/2}$. Recalling the construction of A, $e^{\mathrm{T}}A = e^{\mathrm{T}}B - \delta e^{\mathrm{T}}Q = (1-\delta)e^{\mathrm{T}}$ or $e^{\mathrm{T}}A = e^{\mathrm{T}}B - e^{\mathrm{T}}be^{\mathrm{T}} = (1-\delta)e^{\mathrm{T}}$ for the assumptions (a) or (b), respectively. Since

$$x_s^{\mathrm{T}}A_s(x) = e^{\mathrm{T}}D(x)^{\mathrm{T}}D(x)^{-1}AD(x) = (1-\delta)e^{\mathrm{T}}D(x) = (1-\delta)x_s^{\mathrm{T}}$$

it holds for the assumptions (a) and (b)

$$||A_s(x)||_{x_s} = 1-\delta.$$

Now we consider the assumptions of (a) and (b) separately. Under the assumption (a) it holds $A\hat{x} = (1-\delta)\hat{x}$ while under the assumption (b) we have only $A\hat{x} \leqslant \hat{x}$. For $x = \hat{x}$

$$\begin{aligned} A_s(\hat{x})\hat{x}_s &= D(\hat{x})^{-1}AD(\hat{x})D(\hat{x})^{-1}\hat{x} \\ &= D(\hat{x})^{-1}A\hat{x}. \end{aligned}$$

The last expression is equal to $(1-\delta)\hat{x}_s$ in the case (a) and it is less or equal to $\hat{x}_s$ in the case (b). Thus $||A_s(\hat{x})^{\mathrm{T}}||_{\hat{x}_s} = 1-\delta$ for (a) and $||A_s(\hat{x})^{\mathrm{T}}||_{\hat{x}_s} \leqslant 1$ for (b). Then

$$\begin{aligned} ||A_s(\hat{x})||_2^2 &= r\big(A_s(\hat{x})^{\mathrm{T}}A_s(\hat{x})\big) \leqslant ||A_s(\hat{x})^{\mathrm{T}}A_s(\hat{x})||_{\hat{x}_s} \\ &\leqslant ||A_s(\hat{x})^{\mathrm{T}}||_{\hat{x}_s}||A_s(\hat{x})||_{\hat{x}_s}. \end{aligned}$$

This means that $||A_s(\hat{x})||_2^2 \leqslant (1-\delta)^2$ under the assumption (a) and $||A_s(\hat{x})||_2^2 \leqslant 1-\delta$ under the assumption (b). Let us stress that we have estimated $||A_s(x)||_2$ only for $x = \hat{x}$.

In the rest of the proof, we estimate the norm of $J_s(\hat{x})$. From

$$J(x) = A(I - P(x)) + AP(x)J(x)$$

it follows

$$J_s(x) = A_s(I - P_s(x)) + A_sP_s(x)J_s(x).$$

Since $P_s(x)$ is an orthogonal projection, for any u [14]

$$\begin{aligned} ||J_s(x)u||_2^2 &= ||P_s(x)J_s(x)u||_2^2 + ||(I - P_s(x))J_s(x)u||_2^2 \\ &\leqslant ||A_s(x)||_2^2\left(||(I - P_s(x))u||_2^2 + ||P_s(x)J_s(x)u||_2^2\right). \end{aligned}$$

Using $||A_s(\hat{x})||_2 \leqslant \omega$, $\omega = 1-\delta$ for the case (a) and $\omega = \sqrt{1-\delta}$ in the case (b),

$$||(I - P_s(\hat{x}))J_s(\hat{x})u||_2^2 \leqslant \omega^2||u||_2^2 - (1-\omega^2)||P_s(\hat{x})J_s(\hat{x})u||_2^2,$$

which yields

$$||(I - P_s(\hat{x}))J_s(\hat{x})||_2 \leqslant \omega.$$

Let λ be any eigenvalue of $J_s(\hat{x})$ and v be the corresponding eigenvector. Then $(I - P_s(\hat{x}))v$ is either zero or some eigenvector of $(I - P_s(\hat{x}))J_s(\hat{x})$ corresponding to eigenvalue λ. Thus $|\lambda| \leqslant \omega$. When $J_s(\hat{x}) = D(\hat{x})^{-1}J(\hat{x})D(\hat{x})$, the spectra of the matrices $J_s(\hat{x})$ and $J(\hat{x})$ do not differ. Thus $r(J(\hat{x})) \leqslant \omega$, which finishes the proof. □

Because the spectral radius of a matrix is a continuous function of its elements, there exists a neighborhood of the exact solution $\hat{x}$ such that for any x in it, the spectral radius of $J(x)$ is less than one. Then for any $\epsilon > 0$ there exists a neighborhood of $\hat{x}$ and a matrix norm $||\cdot||_*$ consistent to some vector norm, such that $||J(x)||_* \leqslant \omega + \epsilon$ for any x in the neighborhood of $\hat{x}$. The value ω is less than 1 according to the assumptions of Proposition 2. This results in

Theorem 1. *Algorithm* 1 *is locally convergent when* $T = B$, $s = 1$, $t = 1$ *and* B *contains at least one elementwise positive row. When* $B \geqslant be^{\mathrm{T}}$, $b \geqslant 0$, $||b|| = \delta$, $\delta \in (0, 1)$, *then the asymptotic rate of convergence is at least* $\sqrt{1-\delta}$. *When* $B \geqslant \delta Q$, $\delta \in (0, 1)$, Q *is the Perron projection of* B, *then the asymptotic rate of convergence is at least* $1 - \delta$. *This means, that in each step the approximation error reduces asymptotically by factor* $\sqrt{1-\delta}$ *or* $1 - \delta$, *respectively.*

Example 1. This example shows that the convergence of IAD methods is not guaranteed by the convergence of the basic iteration matrix without some demands upon the number of relaxations on the fine level. It is shown that there exist matrices, for which the IAD algorithm does not converge even locally. Let

$$B = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix}$$

and let the aggregation groups be $\{1\}$ and $\{2, 3\}$. Then the stationary probability vector is $\hat{x} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})^{\mathrm{T}}$ and the spectrum of B is $\sigma(B) = \{1, -\frac{1}{2}, 0\}$. We consider $T = B$ and $s = t = 1$ in Algorithm 1, then the spectrum of $J(\hat{x})$ is $\sigma(J(\hat{x})) = \{-1, 0, 0\}$. Thus the local convergence of IAD Algorithm 1 for $t = s = 1$ and $T = B$ is not ensured for this matrix, because of $r(J(\hat{x})) = 1$. Indeed, if we choose $x^0 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{2})^{\mathrm{T}}$, then the algorithm results in the alternating sequence $x^0, x^1, x^0, x^1, \ldots$, where $x^1 = (\frac{2}{5}, \frac{2}{5}, \frac{1}{5})^{\mathrm{T}}$. But we can also take any vector $\tilde{x}^0 = \hat{x} + \alpha(x^0 - \hat{x})$, for arbitrarily small positive α, i.e. $\tilde{x}^0$ arbitrarilly close to $\hat{x}$, as a starting vector and obtain again the oscillating sequence of approximations. Let us stress that the assumption of Theorem 1 (the positivity of at least one row of B) is not fulfilled and the local convergence condition does not match in this example.

The behavior of the IAD method in Example 1 can be also analyzed using a method described in [5], Algorithm 3. These two algorithms work identically in the case of Example 1. Let us briefly introduce Algorithm 3 here. The set of events is partitioned into two groups and, accordingly, matrix B is (permuted and) partitioned, so that

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}.$$

The starting vector x^0 is choosen and a sequence of the approximations is generated by the formula

$$x^{k+1} = \tilde{\rho}\widetilde{B}x^k,$$

where

$$\widetilde{B} = \begin{pmatrix} 0 & (I - B_{11})^{-1}B_{12} \\ 0 & S \end{pmatrix},$$

where S is the stochastic complement of B_{22} in B, $S = B_{22} + B_{21}(I - B_{11})^{-1}B_{12}$ and $\tilde{\rho}$ is a normalizing factor choosen such that $||x^{k+1}|| = 1$. One can verify that in Example 1

$$\widetilde{B} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

None of the convergence conditions presented in [5, p. 10], primitivity of S or positivity of at least one diagonal element of S, does not match in this example. Note also that Algorithm 1 converges locally for $T = B^2$ and $r(J(\hat{x})) = \frac{1}{2}$ for this case.

Algorithm 3 of [5] can be also viewed as a particular case of the IAD algorithm presented in this paper, where we consider $n - 1$ one-element aggregation groups and one aggregation group containing the rest of events $(N - n + 1)$ and where we take a one-step block Jacobi iteration for smoothing in Step 4.

A similarity of Algorithm 3 of [5] and the IAD method presented in this paper for $T = B$ and $s = t = 1$ can be elucidated from the following lemma.

Lemma 3. *The approximations obtained by the IAD algorithm*, *Algorithm* 1, *fulfill*

$$BP(x^k)x^{k+1} = x^{k+1}.$$

Proof. Recalling Algorithm 1 for $T = B$ and $s = t = 1$, we see that

$$RBS(x^k)z = z.$$

Multiplying by $BS(x^k)$ from left, we have

$$BP(x^k)BS(x^k)z = BS(x^k)z.$$

Since $BS(x^k)z = x^{k+1}$, we obtain validity of the required proposition. □

Lemma 3 says that in fact, the IAD method solves the equation $BP(x^k)x^{k+1} = x^{k+1}$ in each step. We can partition matrix $BP(x^k)$ in accordance with partitioning

given by Algorithm 3 of [5] and get the stochastic complement formulation. Let us suppose two aggregation groups and the corresponding partitioning of matrix $P(x^k)$,

$$P(x^k) = \begin{pmatrix} P(x^k)_1 & 0 \\ 0 & P(x^k)_2 \end{pmatrix}.$$

Then we denote

$$\widetilde{B}_{\mathrm{IAD}}(x^k) = \begin{pmatrix} 0 & (I - B_{11}P(x^k)_1)^{-1}B_{12} \\ 0 & S_{\mathrm{IAD}}(x^k) \end{pmatrix} \begin{pmatrix} 0 \\ P(x^k)_2 \end{pmatrix},$$

where $S_{\mathrm{IAD}}(x^k) = B_{22} + B_{21}P(x^k)_1(I - B_{11}P(x^k)_1)^{-1}B_{12}$. If we take any nonnegative column vector of the size equal to the size of B_{22} and multiply it by $\widetilde{B}_{\mathrm{IAD}}(x^k)$, then we obtain the solution of $BP(x^k)x = x$, i.e. approximation x^{k+1} up to a multiplicative constant. Note that while in Algorithm 3 one iteration with $\widetilde{B}$ yields the next approximation (up to a multiplicative constant) of the exact solution of $Bx = x$, one iteration with $\widetilde{B}_{\mathrm{IAD}}(x^k)$ gives the exact solution of $BP(x^k)x = x$ (up to a multiplicative constant), and thus the next approximation of Algorithm 1. So that, Algorithm 3 of [5] and Algorithm 1 for $T = B$ and $s = t = 1$ are identical if $\widetilde{B}_{\mathrm{IAD}}(x^k)$ does not depend on k and $\widetilde{B} = \widetilde{B}_{\mathrm{IAD}}(x^k)$. When $S_{\mathrm{IAD}}(x)$ does not depend on x and is cyclic, then the IAD method may not converge for some initial approximation. Another case of the absence of convergence is presented in the following example.

Example 2. B is given by

$$B = \begin{pmatrix} 0 & 1 & 0 & 1/2 \\ 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

and the aggregation groups are $\{1, 2\}$ and $\{3, 4\}$. Suppose again $T = B$ and $s = t = 1$. We let to the reader to show the oscillation in the sequence of approximations obtained by Algorithm 1 for some initial vectors x^0.

As shown in [11] there is a lot of free parameters in the IAD algorithm. A crucial role in the convergence issues is played by an interplay between the combinatorial properties of the original stochastic matrix and the basic iteration matrix projected by the complementary aggregation projection. For classical basic iteration schemes it is nicely shown in [3]. From this view point it seems interesting that the schemes with aggregating $\{1\}$ and $\{2, 3\}$ in Example 1 and $\{1, 2\}$, $\{3, 4\}$ in Example 2 offer divergent processes with B as basic iteration matrix while the IAD processes with the same aggregation but different basic iteration matrix may converge rapidly. This is the case when one chooses the block diagonal Jacobi iteration process as the basic. Then the exact solution is returned after at most two iteration sweeps. The rapid convergence is caused by suitable choice of aggregation and the fact that the off-diagonal blocks are special rank-one matrices, see [11].

3. Global convergence of IAD method

We start with an example.

Example 3. We are given a 3×3 stochastic matrix B,

$$B = \begin{pmatrix} 1/12 & 10/12 & 1/12 \\ 1/12 & 1/12 & 1/12 \\ 10/12 & 1/12 & 10/12 \end{pmatrix}.$$

The aggregation groups are $\{1, 2\}$ and $\{3\}$. Assume a positive vector x, $x = \left(\frac{10}{12}, \frac{1}{12}, \frac{1}{12}\right)^{\mathrm{T}}$, and construct the projection matrix $P(x)$, $P(x) = S(x)R$, where

$$S(x) = \begin{pmatrix} x_1/(x_1 + x_2) & 0 \\ x_2/(x_1 + x_2) & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$R = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Let $T = B$ and $s = t = 1$ in Algorithm 1 and let $J(x)$ be the matrix of Lemma 1,

$$J(x) = Z(I - P(x)Z)^{-1}(I - P(x)).$$

Then the spectral radius of $J(x)$ for given x is about 0.0732. But when the first two components of vector x are switched, i.e. we consider vector $y = (\frac{1}{12}, \frac{10}{12}, \frac{1}{12})^{\mathrm{T}}$, then the spectral radius of $J(y)$ increases to the value about 2.1429. Thus we can conclude, that in the proof of the global convergence for the case of $s = t = 1$ and $T = B$ in Algorithm 1, rather than the estimate of the spectrum of $J(\cdot)$ a different technique should be adopted. Another possibility is to consider Algorithm 1 with more general T and with larger t and s.

Lemma 4. *When $B^s \geqslant \delta Q$, $\delta \in (0, 1)$, $T = B$ and $t \geqslant s$, then*

$$J(x) = B^t(I - P(x)V)^{-1}(I - P(x)) = B^{t-s}K(x),$$

where $K(x) = V(I - P(x)V)^{-1}(I - P(x))$ and V is the global core matrix corresponding to B^s and $\hat{x}$.

Proof. The derivation can be done in the same manner as in the proofs of Lemmas 1 and 2. The global core matrix [11] V is here equal to $Z^s + (1 - \delta)Q = B^s - \delta Q$. □

As indicated in Lemma 1, for $T = B$, $t \geqslant s \geqslant 1$, the convergence of Algorithm 1 depends on some norm of the operator J. From Lemmas 2 and 4, we have

$$||K(x)|| \leqslant ||V||||I - P(x)|| + ||V||||P(x)||||K(x)||$$

thus

$$||K(x)|| < \frac{2(1-\delta)}{\delta}$$

because of $||V|| = 1 - \delta$, $||I - P(x)|| < 2$ and $||P(x)|| = 1$ for any positive x.

So that the sufficient condition for the global convergence of IAD can be $(1 >)$ $\delta \geqslant \frac{2}{3}$, i.e. $B^s > \frac{2}{3}Q$. This restrictive condition may often not be matched. In such a case, the value t in $J(x) = B^{t-s}K(x)$ can be estimated to ensure $||J(x)|| < 1$. Having $B^s > 0$, the lower bound of δ is estimated by the minimal element of matrix B^s, $\delta \geqslant \min\{(B^s)_{ij}; i, j = 1, \ldots, N\}$.

Lemma 5. *When* $B^s \geqslant \delta Q$, $\delta \in (0, 1)$, $T = B$, $t \geqslant s$, *then* $||J(x)|| < 1$ *for any* $x > 0$ *if*

$$t \geqslant s\frac{\ln\delta - \ln 2}{\ln(1-\delta)}.$$

Proof. The proposition can be verified by direct computing the exponent t such that $t - s \geqslant sk$ where $(1 - \delta)^k 2(1 - \delta)/\delta < 1$. □

It can be seen, that when $B^s \geqslant \delta Q$, δ is close to 0 and the global convergence is to be ensured, then the exponent t in Step 4 in Algorithm 1 must be of the value approximately $|\frac{\log\delta}{\delta}|$. On the other hand, the power method converges globally with the convergence rate $1 - \delta$ when $B \geqslant \delta Q$. From this point of view, IAD method does not seem to behave better then the power method. But the advantage of IAD methods insits in proper partitioning techniques, as shown e.g. in [2,7,10–13].

4. Conclusions and discussion

In this paper we derived a sufficient condition for local convergence of the IAD method represented by Algorithm 1 when considering $T = B$ and $s = t = 1$, i.e. the smoothing step is represented by one-step power iteration with matrix B. This condition is positivity of at least one row of B. In addition to it, an upper bound of the asymptotic rate of convergence is derived. It is $1 - \delta$ when $B \geqslant \delta Q$ and it is $\sqrt{1-\delta}$ when $B \geqslant be^{\mathrm{T}}$, where Q is the Perron projection of B, $b \geqslant 0$, $||b|| = \delta$, $\delta \in (0, 1)$. It means that the error of the kth approximation asymptotically reduces by factor $(1 - \delta)^k$ or $\sqrt{1-\delta}^k$, respectively.

This new results correct the convergence theorem of [10]. In [10] it was claimed that the spectral radius of the error matrix $J(x)$ (introduced here in Lemma 1) is smaller than one for x equal to the exact solution $\hat{x}$. But we have shown that there exist irreducible primitive matrices B for which $r(J(\hat{x})) = 1$.

Some relations between the IAD method represented by Algorithm 1 and Algorithm 3 analysed in [5] have been shown. We were inspired by the analysis in [5] to

better understanding the divergence issues of IAD methods in some cases. As principal we consider the problem of choosing a reasonable partitioning of a given stochastic matrix into blocks, i.e. the choice of the aggregation groups. This problem remains still open. The establishing a more accurate local convergence sufficient condition based on the sparsity pattern of the original matrix B for the partitioning into more than two groups of events is still also the open question.

The sufficient global convergence condition is derived. When t and s are sufficiently large, we obtain the global convergence of the IAD method. These derivations give rise to quite pessimistic estimates. Some very promissing results are obtained for classes of stochastic matrices possessing some special structures [11]. For problems with such matrices the exact solutions are returned by the IAD algorithms including Algorithm 1 after at most two IAD iteration sweeps. The research in this area will be the subject of our future work.

Acknowledgments

This research was supported by the Grant Agency of Czech Republic under the Contract No. 201/05/0453 and by the Information Society Project No. 1ET400300415. The authors thank to anonymous referee for his valuable comments.

References

[1] A. Berman, R.J. Plemmons, Nonnegative matrices in the mathematical sciences, Classics Appl. Math. 9, SIAM, 1994.
[2] F. Chatelin, W.L. Miranker, Acceleration by aggregation of succesive approximation methods, Linear Algebra Appl. 43 (1982) 17–47.
[3] P.J. Courtois, P. Semal, Block iterative algorithms for stochastic matrices, Linear Algebra Appl. 76 (1986) 59–80.
[4] G.H. Golub, C.F. Van Loan, Matrix Computations, Johns Hopkins University Press, Baltimore, MD, 1996.
[5] I.C.F. Ipsen, S. Kirkland, Convergence Analysis of an Improved PageRank Algorithm, Technical Report, North Carolina State University, 2004.
[6] U.R. Krieger, Numerical solution of large finite Markov chains by algebraic multigrid techniques, in: W.J. Stewart (Ed.), Computations with Markov Chains, Kluwer Academic Publisher, Boston, 1995, pp. 403–424.
[7] A.N. Langville, C.D. Meyer, Updating PageRank using the Group Inverse and Stochastic Complementation, CRSC Technical Report, North Carolina State University, 2002.
[8] C.P. Lee, G.H. Golub, S.A. Zenios, A Fast Two-stage Algorithms for Computing PageRank, SCCM Report, Stanford University, 2003.
[9] J. Mandel, B. Sekerka, A local convergence proof for the iterative aggregation method, Linear Algebra Appl. 51 (1983) 163–172.
[10] I. Marek, P. Mayer, Convergence analysis of an iterartive aggregation/disaggregation method for computing stationary probability vectors of stochastic matrices, Numer. Linear Algebra Appl. 5 (1998) 253–274.

[11] I. Marek, P. Mayer, Convergence theory of some classes of iterative aggregation–disaggregation methods for computing stationary probability vectors of stochastic matrices, Linear Algebra Appl. 363 (2003) 177–200.
[12] P.J. Schweitzer, K.W. Kindle, An iterative aggregation–disaggregation algorithm for solving linear equations, Appl. Math. Comput. 18 (1986) 313–353.
[13] W.J. Stewart, Introduction to the Numerical Solutions of Markov Chains, Princeton University Press, Princeton, 1994.
[14] E.A. Taylor, Úvod do funkcionální analýzy, Academia, Praha, 1973 (in Czech).
[15] R.S. Varga, Matrix Iterative Analysis, 2nd ed., Springer Series in Computational Mathematics, vol. 27, Springer-Verlag, New York, 2000.